Zero : Jurnal Sains, Matematika, dan Terapan
E-ISSN : 2580-5754; P-ISSN : 2580-569X
Volume 9, Number 3, 2025
DOI: 10.30829/zero.v9i3.27088
Page: 842-856

Spatial Grouping of Tornado-Relevant Wind Regimes Areas in Indonesia to Enhance Disaster Risk Mitigation Capacity

[1] Sela Naren Ardelita
Mathematics Department, Universitas Negeri Surabaya, Surabaya, 60231, Indonesia

[2] A'yunin Sofro
Actuarial Science Department, Universitas Negeri Surabaya, Surabaya, 60231, Indonesia

Article Info

Article history:

Accepted, 26 December 2025

Keywords:

Average Linkage;
Brain Storm;
Clustering;
Tornadoes.

ABSTRACT

Tornadoes are major weather hazards in Indonesia, where wind variability is important for assessing disaster risk and supporting energy planning. This study conducts a short-term (one-year) analysis by identify similarities in regional wind speed patterns using a time-series clustering approach, treating monthly average wind speeds in 2024 as proxies for tornado-relevant wind regimes rather than direct tornado occurrence data. Agglomerative hierarchical clustering is integrated with three distance measures—Dynamic Time Warping (DTW), Autocorrelation Function (ACF), and Short Time Series (STS)—and optimized using Brain Storm Optimization (BSO) to determine optimal distance weighting and cluster numbers. The results indicate that DTW provides the best performance, yielding a two-cluster solution with a Silhouette Coefficient of 0.5292. The first cluster exhibits relatively stable wind patterns, while the second shows higher temporal variability. This framework provides a data-driven basis for region-specific wind energy planning and tornado-adaptive infrastructure considerations in Indonesia.

Corresponding Author:

A'yunin Sofro,
Actuarial Science Department
Universitas Negeri Surabaya, Surabaya, Indonesia
Email: ayuninsofro@unesa.ac.id

1. INTRODUCTION

Indonesia is also known as an agricultural country and an archipelago located in a geographical, hydrological, geological, and demographic position that is prone to disasters [1]. One of the atmospheric hazards affecting Indonesia is tornadoes. Tornadoes stand out as some of the most devastating weather events, inflicting substantial harm on structures, essential services, and the natural environment nearby [2].

Numerous public databases have been made available online by different countries, including those in Europe [3], the United States [4], Canada [5], and Japan [6]. Conversely, numerous tornado frequency studies have been documented for various nations throughout Europe [7], Northern Eurasia [8], South America [9], Australia [10], and Asian nations like Japan [11];[12], China [13]; [14], India-Pakistan [15], and the Philippines [16].

In this paper, the definition of a tornado comes from the [17]. A tornado isdescribed as "a spinning column of air that rapidly rotates, extending from the ground up to the bottom of a cumuliform cloud, frequently seen as a funnel cloud." Moreover, this description is expanded to include all waterspouts, regardless of whether they come ashore, consistent with the tornado description by [18].

Tornadoes occur in areas with large temperature gradients [19]. Temperature gradients affect pressure gradients and wind energy [20]. Currently, wind energy has received increasing attention. Although wind energy capacity is relatively low compared to other renewable energy sources such as biomass, hydro, and nuclear power

plants, it is estimated that wind energy will contribute 18% of the world's electricity supply by 2050 [21]. For developing countries such as Indonesia, wind energy is an attractive option for switching from fossil fuels to renewable energy. However, wind energy development in Indonesia is still low. One of the main factors is the relatively low average wind speed, making it difficult to generate electricity on a large scale. Wind speeds are not high enough to build large-diameter wind turbines, as large-scale wind turbines require a minimum wind speed range of between 5 m/s and 7 m/s [22]. This study provides an indirect, wind-regime-based characterization relevant to preliminary tornado risk assessment. Most existing studies address tornado hazards or wind energy potential separately, without explicitly linking wind pattern classification to disaster risk mitigation and resilient wind power development. Furthermore, improved knowledge of tornado occurrence patterns can enhance early warning dissemination, community awareness programs, and emergency response planning, particularly in densely populated and vulnerable areas. This can be done by cluster analysis with similar characteristics.

Cluster analysis groups objects with similar traits while preserving their natural structure, creating meaningful patterns or classifications [23]. Cluster analysis aims to group data objectively into homogeneous groups where the similarities of objects within groups are minimized, and dissimilarities between groups of objects are maximized [24]. The application of cluster analysis is growing with the use of time series data (time series). Time series data is data obtained by observing sequences taken sequentially in time, which has a correlation structure between the values in each time series data [25]. Time series cluster analysis groups objects by their patterns, but the method must account for the dynamic nature of time series data. Calculating the distance between time series objects is one of the main foundations of time series clustering algorithms [26]. Currently, there have been many developments related to the distance measurements used in this study, namely Dynamic Time Warping (DTW) distance, Short Time Series (STS) distance [27], and Autocorrelation Function (ACF) distance [28].

Clustering analysis is divided into two types, namely hierarchical clustering and non-hierarchical clustering. The main difference between the two lies in the initial grouping process. Hierarchical grows clusters gradually, while non-hierarchical fixes cluster count and refines iteratively [29], [30]. According to previous research, the hierarchical method is selected in this study, as this approach is robust to outliers [31]. Hierarchical cluster analysis with agglomerative method consists of single linkage, average linkage, complete linkage, and ward method [32]. But, [33] conducted research on the clustering of districts/cities in South Sulawesi using hierarchical clustering with the average linkage method, and the results showed three clusters: 21 districts/cities, 2 districts/cities, and 1 district/city. In [34] also conducted research on the grouping of secondary crop production using hierarchical clustering by comparing complete linkage and average linkage. The results of the study show that the average linkage method is better because the standard deviation value of average linkage (0.056) is smaller than that of complete linkage. Different from earlier research on grouping wind speeds that depend on just one way to measure distance or a set method for clustering, this research combines multiple time-series distance measures within an optimization-based clustering framework. Grouping wind speed patterns is indirectly relevant to reducing tornado risk, as similar patterns reveal regions with comparable risks and support disaster planning. Grouping recent wind patterns provides insights for resilient wind power planning, disaster preparedness, and integrating risk into energy and infrastructure strategies. However, this analysis represents a short-term snapshot rather than long-term climatology. This research examines the practical and methodological issues involved in grouping short wind speed data at the provincial level in Indonesia. From a methodological angle, the key question is how the methods of Dynamic Time Warping (DTW), Autocorrelation Function (ACF), and Short Time Series (STS) can be integrated into an optimization framework to enhance the clustering of short wind speed data. From a practical standpoint, the study investigates which provinces in Indonesia show similar patterns in wind speeds during 2024 that are important for initial wind energy planning. It should be noted that this analysis is based only on wind speed data, which means it indirectly provides insights on tornado risk and serves as an initial step toward improving disaster risk management and sustainable energy planning.

2. RESEARCH METHOD

2.1 Data Representation and Normalization

2.1.1 Wind Speed

It's crucial to understand that the grouping method hinges on how alike wind speeds are, not on specific records of tornado touchdowns. Because of this, the groups that emerge don't show the definite chance of a tornado happening; instead, they mirror the wind patterns of an area, which are related in theory to how unstable the air is and the risk of strong winds. Wind speed is the speed of air moving horizontally and is influenced by the barometric gradient of the location, the altitude of the location, and the topography [35]. The BMKG classifies wind speed based on the Beaufort scale, which describes wind conditions and visible phenomena in the surrounding environment. There are 12 categories of wind speed according to the Beaufort scale [36]. Information related to the categorization of wind types according to the Beaufort scale can be seen as follows:

Table 1. Wind Speed on Beaufort Scale

Wind Speed (m/s)	Wind Category
0 – 0,2	Calm
0,3 – 1,5	Light air
1,6 – 3,3	Light breeze
3,4 – 5,4	Soft breeze
5,5 – 7,9	Moderate breeze
8,0 – 10,7	Strong breeze
13,9 – 17,1	Light windstorm
17,2 – 20,7	Windstorm
20,8 – 24,4	Strong windstorm
24,5 – 28,4	Storm
28,5 – 32,6	Violent storm
≥ 32,7	Cyclone

Systematic and continuous wind speed measurements enable the identification of seasonal patterns and trends. Monitoring wind speed through time series analysis is essential for predicting future events or evaluating the continuity of wind supply in a particular region [37].

2.1.2 Preprocessing Data

A completeness check confirmed that the monthly wind speed dataset for all 34 provinces contains no missing observations over the January–December 2024 period. Consequently, no data imputation or correction was performed prior to normalization. Detailed information on provinces, station codes, geographic coordinates, elevation, and valid observation periods is provided in Appendix A.

Normalization data processes common data distribution with the aim of normalization. One of the best methods of normalization is Min-Max normalization. Min-Max normalization maps the value from each variable into the same range. Normalization can be calculated by Equation (1) with x as data per column, min as the minimum value of data per column, and max as the maximum value of data per column [38]. The Min-Max normalization formula is written as follows:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

Where x denotes the original observation, and $x_{\min}$ and $x_{\max}$ represent the minimum and maximum values within each variable.

It is acknowledged that the resulting time series consist of only 12 monthly observations, which is extremely short for robust time-series clustering. Consequently, DTW, STS, and ACF distances were selected due to their suitability for short-length, pattern-based, and non-stationary time series, rather than for long-term temporal modeling.

2.2 Distance Matrix Calculation

2.2.1 Dynamic Time Warping (DTW) Distance

The first distance is Dynamic Time Warping (DTW) distance is an algorithm used to find the distance between two time series that have the same or different lengths. With the Dynamic Time Warping (DTW) algorithm, the distance is determined by finding the optimal warping path between two time series, so that the warping path values and the distance between the two time series are obtained. Dynamic Time Warping (DTW) uses dynamic programming by searching for all possible paths and selecting one path with the minimum cumulative distance between two time series data using a distance matrix. Given two time series data $X = x1, x2, ..., xi, ..., xn$ with length n and $Y = y1, y2, ..., yj, ..., ym$ with length m. Next, create a matrix D with size $n \times m$. The element (i,j) of matrix D is the difference between xi and yj, which can be formulated with the following equation:

$$d_{i,j} = |x_i - y_j| \tag{2}$$

which is $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$. The implementation of the D matrix is as follows:

$$D = \begin{bmatrix} d_{1,1} & d_{1,2} & d_{1,3} & \dots & d_{1,j} \\ d_{2,1} & d_{2,2} & d_{2,3} & \dots & d_{2,j} \\ d_{3,1} & d_{3,2} & d_{3,3} & \dots & d_{3,j} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{i,1} & d_{i,2} & d_{i,3} & \dots & d_{i,j} \end{bmatrix} \quad (3)$$

After obtaining the distance, it is then added to a minimum of 3 adjacent elements $\{d_{(i-1)(j-1)}, d_{(i-1)j}, d_{i(j-1)}\}$ where $0 < i \leq n$ and $0 < j \leq m$, thus forming matrix E. Equation (4) can be defined for element (i,j) in matrix E as follows:

$$E_{i,j} = d_{ij} + \{d_{(i-1)(j-1)}, d_{(i-1)j}, d_{i(j-1)}\} E_{ij} \quad (4)$$

The iterative equation for the cumulative cost matrix within the Dynamic Time Warping (DTW) algorithm is presented in Equation (4). Specifically, each element $E_{i,j}$ represents the minimum cumulative distance required to align the subsequence $(x_1, \dots, x_i)$ with $(y_1, \dots, y_j)$. The local distance $d_{i,j} = | x_i - y_j |$measures the instantaneous dissimilarity between the two time series at positions $i$and j. This local distance is then added to the minimum cumulative cost among the three admissible predecessor cells, namely $E_{i-1,j-1}$(diagonal move), $E_{i-1,j}$(vertical move), and $E_{i,j-1}$(horizontal move). These shifts align with the operations of matching, adding, and removing within the warping route, in that order. Dynamic Time Warping (DTW) works by propagating the lowest cumulative cost through the matrix to determine the optimal warping path, minimizing alignment differences between two time series. The implementation of the E matrix is as follows:

$$E = \begin{bmatrix} E_{1,1} & E_{1,2} & E_{1,3} & \dots & E_{1,j} \\ E_{2,1} & E_{2,2} & E_{2,3} & \dots & E_{2,j} \\ E_{3,1} & E_{3,2} & E_{3,3} & \dots & E_{3,j} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ E_{i,1} & E_{i,2} & E_{i,3} & \dots & E_{i,j} \end{bmatrix} \quad (5)$$

After the E matrix is formed, the DTW distance between two time series X and Y can be calculated using Equation (6) below [39]:

$$d_{DTW}(\mathrm{X}, \mathrm{Y}) = \min_{\forall w \in p} \{ \sum_{i,j=1}^{k} E_{i,j} \} \quad (6)$$

Where, p is a set of all possible warping paths, Ei,j is elements (i,j) in matrix E, and k is the length of the warping path.

Equation (6) represents the final Dynamic Time Warping (DTW) distance between two time series $X$and Y, obtained as the minimum cumulative cost along the optimal warping path in the accumulated cost matrix E. The derived DTW distance is a scalar value that is not negative and measures the total dissimilarity between two time series once non-linear temporal alignment has been taken into account. Concerning monthly wind speed data, a lower DTW distance implies that two provinces have very comparable wind speed trends over time, even when peak events orseasonal changes are moved in time. In contrast, greater DTW distances show significant variations in wind speed dynamics, such as opposing seasonal variability or strength. As a result, this DTW distance matrix acts as the main data for the clustering procedure, enabling provinces with similar wind behavior features to be clustered together using the agglomerative hierarchical clustering method.

2.2.2 Autocorrelation Function (ACF) Distance

Galeano and Pella [40] conducted research on the relationship between time series data using the Autocorrelation function (ACF) approach. Suppose we are given two time series data X_t and Y_t, where t is the length of the time series. The autocorrelation function (ACF) distance consists of autocorrelation vectors from X_t and Y_t, namely $\hat{\rho}_{X_t} = (\hat{\rho}_{1,X}, \hat{\rho}_{2,X}, \dots, \hat{\rho}_{k,X})t$ and $\hat{\rho}_{Y_t} = (\hat{\rho}_{1,Y}, \hat{\rho}_{2,Y}, \dots, \hat{\rho}_{k,Y})t$ respectively represent the estimated autocorrelation vectors from lag 1 to lag k where $\hat{\rho}_{i,X_t} \approx 0$ and $\hat{\rho}_{i,Y_t} \approx 0$ for $i > k$. Suppose the time series data $Z = Z_1, Z_2, \dots, Z_n$, then the ACF value can be calculated using equation (7).

$$\rho_k = \frac{\sum_{i=1}^{n=k}(X_i - \bar{X})(X_{t+k} - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \tag{7}$$

Equation (7) computes the sample autocorrelation coefficient $\rho_k$at lag k, which measures the linear dependence between observations of a time series separated by k time steps. In this equation, X_i denotes the wind speed observation at time i, X_{i+k} represents the observation at lag k, $\bar{X}$ is the mean of the time series, and n is the total length of the series. The numerator captures the covariance between lagged observations, while the denominator normalizes the value by the total variance of the series, ensuring that ρ_k lies within the range $[-1,1]$. In this research, a collection of autocorrelation values at various time intervals are arranged to form autocorrelation vectors, which describe the pattern of how monthly wind speeds depend on each other over time. Following this, the resemblance between two regions is determined by assessing the distance between each of their unique autocorrelation vectors, giving another way to understand wind speed activity that goes beyond only comparing the strength of the wind.

After obtaining the autocorrelation vectors, the ACF distance between two time series data X_t and Y_t can be calculated using the following equation:

$$d_{ACF}(X_t, Y_t) = \sqrt{(\hat{\rho}_{X_t} - \hat{\rho}_{Y_t})'\Omega(\hat{\rho}_{X_t} - \hat{\rho}_{Y_t})} \tag{8}$$

Equation (8) defines the Autocorrelation Function (ACF) distance between two time series $X_t$and $Y_t$based on the Euclidean distance between their respective autocorrelation vectors. In this formulation, $\hat{\rho}_{X_t}$and $\hat{\rho}_{Y_t}$ denote the estimated autocorrelation vectors of the two wind speed time series, constructed from autocorrelation coefficients at selected time lags. The matrix Ω represents a weighting matrix that determines the relative contribution of each lag; in this study, Ω is set as an identity matrix, implying equal importance across all lags. The ACF distance that arises gauges the resemblance of how time affects different regions; this allows the grouping procedure to detect shared traits in wind duration and yearly patterns, instead of just focusing on the actual strength of the wind.

Where $d_{ACF}(A, B)$ is the distance between autocorrelation A and B, and Ω is the weighting matrix. However, the ACF distance in this study does not have a weight, so Ω is an identity matrix. Thus, from equation (8), the ACF distance becomes the Euclidean distance between the autocorrelation function estimates with the following equation:

$$d_{ACF}(X_t, Y_t) = \sqrt{\sum_{i=1}^{k}(\hat{\rho}_{X_t} - \hat{\rho}_{Y_t})^2} \tag{9}$$

2.2.3 Short Time Series (STS) Distance

Short Time Series (STS) distance was introduced by Möller-Levet et al. [41] and is used to measure the similarity of DNA microarray time series data. Microarray is a pattern obtained from the analysis of the function and expression of a large number of genes simultaneously in a single experiment. Möller aimed to determine a distance that could capture differences in form, determined by relative expressive changes and corresponding temporal information. Suppose there are two time series data sets $X = \{x_0, x_1, \dots, x_{N-1}\}$ and $Y = \{y_0, y_1, \dots, y_{N-1}\}$, the STS distance is defined in Equation (10) as follows:

$$d_{STS}(X, Y) = \sqrt{\sum_{k=0}^{N-1}\left(\frac{y_{k+1} - y_k}{t_{k+1} - t_k} - \frac{x_{k+1} - x_k}{t_{k+1} - t_k}\right)^2} \tag{10}$$

where t_k is the time of each point in the X and Y data.

2.3 Construction and Normalization of Distance Matrices

In order for each distance matrix to be on a comparable scale, normalization to the interval is performed in [0,1] with method min-max normalization:

$$P_{i,j} = \frac{D_{i,j} - \min(D)}{max(D) - \min(D)} \tag{11}$$

This results in three normalized distance matrices:

a. P_1 = DTW normalized matrix
b. P_2 = ACF normalized matrix
c. P_3 = STS normalized matrix

The diagonal values of each matrix are set to 0 to ensure $d_{(i,j)} = 0$.

2.4 Agglomerative Hierarchical Clustering (Average Linkage)

Agglomerative hierarchical clustering has been the dominant approach to constructing embedded classification schemes. It is our aim to direct the reader's attention to practical methods that are both effective and efficient [42]. Agglomerative hierarchical clustering was then performed on D_{temp} using the Average Linkage method:

$$d(C_a, C_b) = \frac{1}{|C_a|\,|C_b|} \sum_{i \in C_a} \sum_{j \in C_b} D_{i,j} \tag{12}$$

2.5 Determining the k range using Elbow (WSS)

The Elbow method is used only to determine the range of candidate k values, not to select the final k value. Temporary clustering is performed using a combined distance with balanced weights:

$$D_{temp} = \frac{P_1 + P_2 + P_3}{3} \tag{13}$$

For each $k \in \{2, \ldots, 8\}$ calculated Within-Cluster Sum of Squares:

$$WSS = \sum_{c=1}^{k} \sum_{i \in C_c} ||X_i - X_c||^2 \tag{14}$$

Given that the cluster count, represented by k, is a discrete value, a second-order finite difference method is employed to estimate the curvature of the WSS curve. This discrete approximation involves using the second difference of WSS, which is defined as:

$$\Delta^2 WSS(k) = WSS(k+1) - 2WSS(k) + WSS(k-1), k = 3, \ldots, K-1 \tag{15}$$

The elbow index is determined using the second derivative of WSS:

$$k_{elbow} = arg \max_k \left| \frac{\Delta^2 WSS(k)}{dk^2} \right| \tag{16}$$

Rather than fixing a single value of k, the elbow point is used to define a candidate search interval for the optimization procedure. Then, the k range is determined:

$$k_{min} = \max(2, k_{elbow} - 1) \tag{17}$$

$$k_{max} = k_{elbow} + 2 \tag{18}$$

The initial range $k \in \{2, \ldots, 8\}$is selected to ensure sufficient cluster interpretability given the limited number of provinces and the short length of the time series. Range k_{min}, k_{max} is used as the search space for the value k in BSO.

2.6 Brain Storm Optimization

To improve the optimality of the AHC method in this study, an optimization approach is needed that can adjust the distance and cluster search space parameters based on the characteristics of time series data. Brain Storm Optimization (BSO), inspired by the collaborative creativity process of humans in searching for the best solutions, has been used in a number of studies to optimize cluster structures in hierarchical bottom-up methods, including agglomerative clustering. In [43], hierarchical clustering is applied. Meanwhile, the impact on the performance of the creation operator is discussed in depth. A new BSO with hierarchical clustering is then proposed. Agglomerative hierarchical clustering not only does not require a predetermined number of clusters, but also helps the creation operator to improve search performance in exploration and exploitation. The convergence curve and statistical results show that the proposed method can quickly identify regions with high-quality solutions in the search space and obtain satisfactory solutions. On contrast, in [44], the human

brainstorming process is modeled, based on which two versions of the Brain Storm Optimization (BSO) algorithm are introduced. Simulation results show that both BSO algorithms perform well on ten benchmark functions, proving the effectiveness and usefulness of the proposed BSO algorithm. Expanding on these investigations, this article puts forward an improved BSO-based approach using agglomerative hierarchical clustering in order to fine-tune both the significance of distance measurements and the quantity of clusters for the examination of wind speed data over time. Agglomerative hierarchical clustering is especially well-suited for this application because it does not need a predetermined cluster count and offers an adaptable hierarchical arrangement that facilitates the creation of fresh potential solutions while ideas are being developed. Brain storm optimization algorithm (BSO), which is inspired by brain storm process of human, has been adopted as an efficient optimizer for various complex problems [45]. The Brain Storm Optimization algorithm operates according to the following procedure [42]:

Input:

Candidate initial population solutions

Three normalized distance matrices

Candidate cluster range $k \in [k_{min}, k_{max}]$

Maximum iteration parameter $MaxIter = 100$

Solution Representation:

$s = (w_1, w_2, w_3, k)$

where w_1, w_2, w_3 denote the weights associated with the DTW, ACF, and STS distance matrices, respectively, and k reperesnts the number of cluster. The constraints imposed are $w_1 + w_2 + w_3 = 1$ and $w_j \geq 0$ for all j, with $k \in [k_{min}, k_{max}]$

Output:

Optimal distance weight (w_1, w_2, w_3)

Number of best clusters k_{best} in the candidate range k

Algorithm steps:

1. Initialization:
   a. Initialize the population with raw distance weights and k values sampled randomly from the candidate range. The initial distance weights are generated using a uniform random distribution and subsequently normalized to satisfy the weight constraints, while the cluster number k is sampled from a discrete uniform distribution over $[k_{\min}, k_{\max}]$.
   b. Select a number of initial centroids k as the initial solution representation for the brainstorming stage.
2. Assignment:
   a. For each solution in the population, calculate the combined distance matrix based on the weights:

$$D_{combined} = w_1 P_1 + w_2 P_2 + w_3 P_3 \tag{18}$$

   b. Assign each object to the nearest cluster using the combined distance.
   c. Form a temporary cluster hierarchy using the Agglomerative Hierarchical Clustering (Average Linkage) method.
3. Updating:
   a. Update distance weights through brainstorming mechanisms (local recombination and global mutation). To utilize local search areas, solution vectors inside the same cluster core are combined to conduct local recombination.
   b. Evaluate all candidates for new solutions using the Silhouette Coefficient as a fitness function. The fitness of each candidate solution is defined as the global average Silhouette Coefficient computed across all provinces, and the optimization objective is to maximize this value.
   c. Maintain solutions with the highest fitness to shape the next generation of the population.
4. Termination: End the process when the termination criteria are met (unchanged between iterations). The algorithm terminates when either the maximum number of iterations (MaxIter) is reached or no improvement in the best fitness value is observed over successive iterations; otherwise, return to Step 2 and repeat the steps.

Robustness Verification of Linkage Criteria

Following the acquisition of the conclusive clustering result derived from the BSO enhanced agglomerative hierarchical clustering employing average linkage, a verification of robustness was executed through the implementation of Ward and complete linkage methodologies, utilizing an identical Dynamic Time Warping distance metric and the most favorable quantity of clusters. This comparative analysis is designed to evaluate the consistency of the determined cluster arrangement in relation to varying linkage standards.

2.7 Silhouette Coefficient

The silhouette coefficient is a measure used to assess the quality and strength of clusters, particularly to measure how accurately an object is placed in a particular cluster in time series clustering. This method combines two main concepts of cohesion and separation, where cohesion refers to the measurement of the proximity between an object and other objects in a cluster, while separation refers to the measurement of how far an object is from objects in other clusters [46].

The Silhouette Coefficient calculation process begins by calculating a(i), which is the average distance between object i and other objects in the same cluster, also known as cohesion. a(i) can be written in the following equation:

$$a(i) = \frac{1}{|A| - 1} \sum_{j \in A, j \neq i} d(i,j) \tag{19}$$

where A is a cluster, i and j are objects in cluster A, and d(i,j) is the distance between objects i and j [47].

The next step is to calculate b(i), which is the average value of the distance of data i to other clusters, and select the smallest value as the separation. Where C is another cluster different from A, so that it can be expressed in another equation as follows:

$$b(i) = \min_{C \neq A} \left(\frac{1}{|C|} \sum_{j \in C} d(i,j) \right) \tag{20}$$

The Silhouette Coefficient value for object i can be calculated using the following equation:

$$s(i) = \frac{b(i) - a(i)}{\max\big(a(i), b(i)\big)} \tag{21}$$

The following are the criteria for measuring the Silhouette Coefficient [48]:

Table 2. Silhouette Coefficient Criteria

Silhouette Coefficient Value	Cluster Criteria
0,71 - 1,00	Strong
0,51 - 0,70	Moderate
0,26 - 0,50	Weak
0,00 - 0,25	Bad

It's important to recognize that the Silhouette Coefficient, when utilized for internal validation, could be affected by the presence of outliers and how distances are distributed, most notably when using hierarchical clustering techniques. To deal with these possible problems, this research incorporates Dynamic Time Warping (DTW) to lessen the effect of differences in timing and uses Brain Storm Optimization (BSO) for the purpose of automatically improving distance weighting and cluster arrangement. Moreover, checking how well it works with different linkage standards helps to prove that the discovered clustering structure is consistent.

2.8 Data Source

The dataset in this study, includes monthly average wind speeds for 34 provinces in Indonesia during the period from January 2024 to December 2024 [49]. Three different observation stations were selected for each province based on altitude and geographical location to ensure that the data obtained represented the characteristics of the province. The monthly wind speed data from the three stations in a province was then averaged to produce a single value used as a representation of the province's monthly wind speed, was obtained from the Meteorology, Climatology, and Geophysics Agency (BMKG) of Indonesia. To accurately represent the various wind conditions present in each province, encompassing coastal, lowland, and elevated areas, three different monitoring locations were chosen, contingent upon the accessibility and reliability of their data. The monthly wind speed measurements gathered from these locations were then used to calculate an average, yielding a province-level metric designed to showcase the typical wind patterns of the region, as opposed to any unusual or isolated instances. The data is publicly available and can be accessed through the official publications databases of BMKG Indonesia. Even though the data spans merely the year 2024, the assessment seeks to identify present-day wind pattern resemblances among provinces that are pertinent to both infrastructure development in progress

and evaluations of potential disaster threats. Rather than determining enduring climate-based wind behaviors, the study intends to deliver a timely overview of immediate wind activity. This overview can be used to guide approaches to energy development and risk reduction that are planned for the near or intermediate future.

The entire data processing and analysis were conducted using RStudio, a statistical computing and visualization software environment. To begin the research process, data preprocessing was conducted, which included data cleaning and data standardization. After preprocessing, DTW, ACF, and STS distance calculations were performed, and the distance calculation matrix was normalized. Then, the best distance was produced by Agglomerative Average Linkage using the Silhouette Coefficient, then range number of clusters was identified using the Elbow method criteria with the BSO. Subsequently, cluster analysis was performed using the and Agglomerative Average Linkage methods and Brain Storm methods, which produced the best weights and the best number of clusters. Finally, the quality of the resulting clusters was validated through the Silhouette Coefficient to assess their accuracy and determine the most appropriate clustering method. The following section explains the results and analysis in detail. RStudio was employed for all computational analyses and data processing, which offered a versatile and repeatable setting for statistical calculations and examining data over time. The dtwclust package was utilized to determine Dynamic Time Warping distances, and the stats package was used to carry out hierarchical clustering that combines clusters using average, Ward, and complete linkage methods. The cluster package's silhouette function was applied to assess how well the clusters were formed, and tidyverse and readxl were used to help with managing and getting the data ready. A completely unique R script was written to implement Brain Storm Optimization, which specifically defines the solution vector (cluster number and distance weights), makes sure normalization rules are followed, and uses the global silhouette coefficient to judge fitness. To maintain clarity and make sure the results can be reproduced, the default settings of the packages used were kept, unless the script stated otherwise

3. RESULT AND ANALYSIS

The preliminary phase of this study involves data collection and preprocessing. The initial phase of working with wind speed information involved checking the data to find any absent or illogical entries. The ultimate monthly wind speed dataset employed for clustering did not exhibit any absent data points; hence, there was no necessity to implement any methods for either data imputation or removal throughout the analysis. Even though things like mistakes in how wind speed was measured at particular spots or differences in the environment from place to place can have an impact on the specific wind speed numbers, the way we grouped the data focused more on how the wind changed over time relative to other locations; this approach lessened the effect of minor, local data problems. Next, use the Elbow Method criterion to identify the most suitable range of cluster. Figure 1 below presents the results of the cluster number identification.

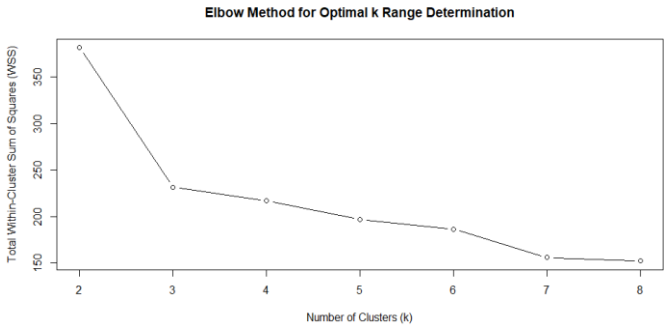

Figure 1. Elbow Method Plot

According to Figure 1, indicates a clear elbow at k = 2, while the second derivative of WSS suggests a plausible range of k = 2–5. This interval was therefore used as the search domain for the BSO optimization stage.

Next, the range cluster is then entered into the BSO implementation in AHC to produce the best coefficient value, best distance and the best number of clusters as well as the best distance weight after several iterations. Table 3 below shows the best distance using Silhouette Coefficient.

Table 3. Result Distance with Silhouette Coefficient

Distance	Silhouette Coefficient
Dyanmic Time Warping	0.5300564
Autocorrelation Function	0.3228557
Short Time Series	0.4145011

Table 3 shows that distance metrics for short time-series data: Dynamic Time Warping (DTW), Auto-Correlation Function (ACF), and Short Time Series (STS) were evaluated using the Agglomerative Hierarchical Clustering (AHC) method combined with the Silhouette coefficient. The evaluation results show that Dynamic Time Warping obtained the highest Silhouette value, namely 0.5300564. Meanwhile, the Autocorrelation Function metric produced a value of 0.3228557 and Short Time Sseries produced a value of 0.4145011. These values confirm that Dynamic Time Warping has the best ability to form cohesive clusters (intra-cluster compactness) while being optimally separated from other clusters (inter-cluster separation). Therefore, Dynamic Time Warping is identified as the best distance metric among the alternatives tested.

At the end of the optimization process, the best configuration was obtained with best Silhouette = 0.5292 and best k = 2. The combination of final distance weights produced by BSO shows the dominance of DTW contribution, with DTW weight value = 0.9983, ACF weight ≈ 1×10^{-4}, and STS weight = 0.0017. These findings indicate that DTW is the main factor in the best clustering structure found by BSO, while ACF and STS only make minor contributions.

After BSO found the best weight and k, the final clustering was built using the Agglomerative Hierarchical Clustering (AHC) method with the Average Linkage scheme based on the weighted combined distance matrix resulting from BSO. The resulting clusters consisted of two groups of provincial clusters, which represented the most optimal pattern of monthly wind speed similarity according to the algorithmic evaluation performed. Cluster-level descriptive statistics in Table 4 indicate that Cluster 1 exhibits higher mean wind speeds with lower variability compared to Cluster 2, which shows lower averages and wider dispersion. These differences quantitatively support the interpretation of more stable wind regimes in Cluster 1.

Table 4. Cluster-Level Descriptive Statistics

Cluster	Mean	Standart Dev.	Min	Max
1	2.01296	0.42749	1.24444	3.42528
2	1.27569	0.34424	0.50038	2.51612

The results from the AHC-BSO clustering suggest that Cluster 1 exhibits a wind pattern that is comparatively stronger and more consistent, potentially making it suitable for evaluations of large renewable energy projects under different conditions, especially when paired with examinations of technical practicality, financial viability, and ecological impact. This pattern is supported by the characteristics of several provinces included in this cluster, such as South Sulawesi, Bali, and South Kalimantan. All of them of which have wind power plants. For example, South Sulawesi has the Sidrap wind power plant with a capacity of 70 MW and 30 Wind Turbine Generators (WTGs) [50]. Meanwhile, Bali has the Nusa Penida wind power plant with 80 kW with 9 units [51]. South Kalimantan has Tanah Laut wind power plant project with a capacity of 70 MW and 10 MW/10 MWh battery storage in the Kalimantan Electricity Grid. The Tanah Laut wind power plant project reaffirms Indonesia's commitment to achieving net zero emissions by 2060 [52]. This condition reinforces the interpretation that the provinces in Cluster 1 are suitable locations for wind power plants supported by advanced electricity grid integration and energy storage systems. In addition, from a scenario-based perspective, robust wind infrastructure in these regions could potentially be designed to support broader early warning and resilience strategies. However, such applications would require further integration with severe weather observations, tornado occurrence records, and convective environment indicators. The implementation of predictive wind monitoring from turbine networks can also contribute to the identification of local tornado risks and improve adaptive disaster mitigation planning.

On the contrary, Cluster 2 describes areas with lower or more fluctuating wind speeds, require strategic interventions through hybrid renewable energy systems. This pattern is supported by the characteristics of several provinces included in this cluster, such as West Kalimantan and North Sulawesi, both of which have hybrid renewable energy systems. For example, West Kalimantan has undergone an energy transition in Temajok Village through the implementation of a hybrid solar-diesel power plant system, with solar power plant performance in 2024 showing stability with a total production of 200,528 kWh, a significant increase compared to 37,497 kWh at the end of 2023 [53]. Meanwhile, North Sulawesi has Kawaluso island, is one of the outhermost islands in northern Sulawesi Electricty in Kawaluso island depends from Diesel energy and it has 12 hour operation. with a distance of 68 KM from the city of Tahuna With the cost of generation up to Rp. 10.360/kWh, its important to find the alternative energy to make lower cost of generation than diesel generator [54]. This condition reinforces the interpretation that provinces in Cluster 2 are better prepared for energy diversification and optimal site selection to improve wind capture efficiency. These examples are presented for contextual illustration only and do not constitute a systematic statistical validation of the relationship between cluster membership and existing wind energy infrastructure. From an analytical perspective, the clustering outcomes provide a structured summary of regional wind variability patterns, which may inform subsequent applied studies on energy planning and hazard-related infrastructure design. It should be emphasized that the clustering is derived solely from wind speed similarity patterns. Therefore, any implication related to tornado hazards should be

interpreted as a hypothesis rather than a confirmed finding. The identified clusters might serve as one component in a broader multi-hazard framework, conditional on further integration with severe convective weather indicators and documented tornado occurrence data.

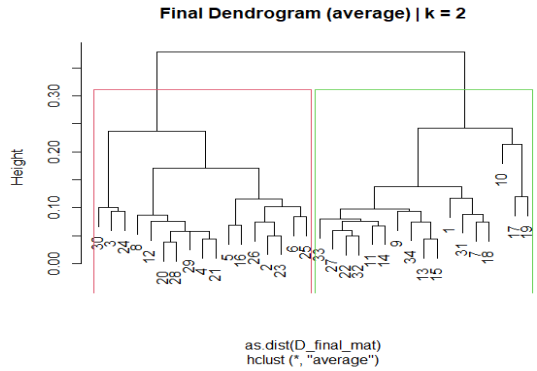

Figure 2. Dendrogram Result of AHC-BSO

The dendrogram above shows the result of Agglomerative Average Linkage clustering with Brain Storm Optimization generated in RStudio, with Cluster 1 marked in green and Cluster 2 in red. The result AHC-BSO Clustering shown in Table 5 below.

Table 5. Result AHC-BSO Clustering

Cluster	Provinces	Number of Members
Cluster 1	Aceh, Bengkulu, Kepulauan Bangka Belitung, Kepulauan Riau, DKI Jakarta, Jawa Tengah, DI Yogyakarta, Jawa Timur, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Selatan, Sulawesi Selatan, Maluku, Maluku Utara, Papua Barat, and Papua.	17
Cluster 2	Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Lampung, Jawa Barat, Banten, Kalimantan Barat, Kalimantan Tengah, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat.	17

To assess the robustness of the identified clustering structure, a comparative analysis using alternative linkage criteria was conducted. Using the same Dynamic Time Warping (DTW) distance and the optimal number of clusters obtained from the BSO optimization, agglomerative hierarchical clustering was re-applied with Ward and complete linkage methods. The results show that the average linkage (BSO-based) and Ward linkage yield identical Silhouette Coefficient values of 0.5301, while the complete linkage produces a slightly lower value of 0.5246. This consistency across different linkage methods indicates that the two-cluster solution is stable and not sensitive to the choice of linkage criterion. Therefore, the selected clustering configuration reliably captures the underlying wind speed pattern structure across Indonesian provinces. It should be emphasized that all hazard- and policy-related interpretations in this study are scenario-based and exploratory, derived from wind speed similarity patterns rather than from direct tornado occurrence data or severe convective environment indicators. Further validation shows that the average pairwise distance within clusters is substantially lower than between clusters, confirming good intra-cluster compactness in Appendix B. In addition, province-level silhouette in Figure 3 below scores are predominantly positive across both clusters, indicating consistent membership assignment.

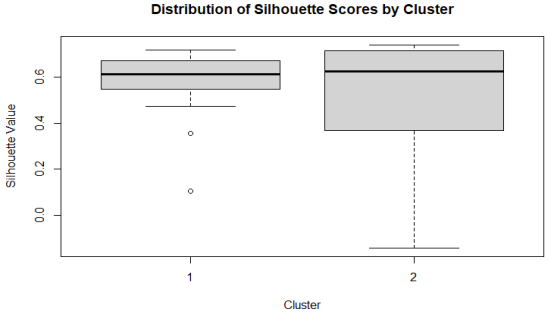

Figure 3. Distribution of Silhouette Score by Cluster

A leave-one-month-out stability analysis in Appendix C was performed to assess the robustness of the clustering structure against temporal perturbations. In this analysis, one month was sequentially removed from the monthly wind speed time series, and the clustering procedure was re-applied using the same optimal configuration obtained from the AHC–BSO framework. The agreement rate between the perturbed clustering results and the baseline clustering was then calculated for each iteration. The results show that the agreement rate ranges from 0.7647 to 1.0000 across all months, with an average stability value of 0.9314. Several months (B24, C24, D24, F24, and J24) exhibit perfect agreement, indicating no change in cluster membership. Even in the least stable cases, more than 76% of provinces remain consistently assigned to the same clusters. These findings demonstrate that the identified two-cluster solution is highly stable and not dominated by any single monthly observation, confirming the robustness of the wind speed clustering results.

It should be noted that the results presented in this section are subject to several limitations. The analysis relies on monthly wind speed data from a single year (2024), which may not fully represent long-term wind variability. In addition, the relatively short time series and the use of a single variable (wind speed) constrain the interpretation of atmospheric dynamics. Moreover, no direct tornado occurrence data or severe weather indices are incorporated; therefore, any discussion related to tornado risk should be interpreted as inferential and conditional on further integration with multi-variable meteorological and hazard datasets.

Based on these findings, region-specific policy recommendations can be formulated according to the wind speed characteristics of each cluster. Cluster 1 provinces, which exhibit relatively higher and more stable wind profiles, can prioritize the development of large-scale renewable energy projects, particularly wind farm deployment supported by advanced grid integration and energy storage systems. Meanwhile, Cluster 2 provinces, characterized by lower or more fluctuating wind speeds, require strategic interventions through hybrid renewable energy systems, localized energy diversification, and optimization of site selection to improve energy capture efficiency. The regional wind behaviors that have been discovered offer a useful substitute for guiding tornado-vulnerable infrastructure design and disaster readiness approaches, even if the clusters are created using wind-speed similarities instead of precise tornado occurrence data.

4. CONCLUSION

Based on wind speed clustering results across Indonesian provinces in 2024, the BSO-guided framework identified Dynamic Time Warping (DTW) as the most representative distance metric, yielding the highest Silhouette Coefficient (0.5301) and an optimal two-cluster configuration (global Silhouette = 0.5292). Beyond supporting region-specific wind energy planning, these clusters provide insights relevant to infrastructure resilience and wind-related hazard assessment, conditional on further integration with severe weather indicators and explicit tornado occurrence data. Provinces in Cluster 1, characterized by relatively stable and higher wind speeds, are suitable for large-scale wind power development under controlled engineering standards, where predictable wind regimes allow for optimized turbine design and grid integration with lower exposure to abrupt wind hazards. In contrast, Cluster 2 provinces, which exhibit more fluctuating wind patterns associated with convective variability, require tornado-adaptive infrastructure strategies, including reinforced turbine structures, flexible foundation systems, and spatial zoning that avoids high-exposure corridors.

From a policy perspective, the clustering outcomes support risk-informed disaster management, enabling authorities to prioritize early-warning integration, infrastructure strengthening, and land-use regulation in tornado-prone regions. This study thus contributes a data-driven bridge between wind climatology, disaster risk mitigation, and renewable energy infrastructure planning, offering practical guidance for resilient development in Indonesia. From an applied mathematics perspective, future research may extend this framework to multivariate spatio-temporal clustering by incorporating additional atmospheric variables and spatial dependencies. Further work is also needed to establish theoretical guarantees or bounds for BSO-based distance weighting and to compare its performance with alternative optimization schemes, such as genetic algorithms or Bayesian model selection approaches.

5. REFERENCES

[1] Siswanto, & Supari. "Identifying Precursor Condition for Puting Beliung Event in Pangkalpinang." *Widyariset,* 15(3), 599–610, 2012. [Online]. Available: http://dx.doi.org/10.14203/widyariset.15.3.2012.599-610

[2] V. A. Johnson, K. E. Klockow-McClain, R. A. Peppler, and A. M. Person, "Tornado climatology and risk perception in central Oklahoma," *Weather, Climate, and Society*, vol. 13, pp. 743–751, 2021. [Online]. Available: 10.1175/WCAS-D-20-0137.1.

[3] "European Severe Weather Database, or ESWD", *Dotzek et al., 2009.* Accessed: Dec, 16, 2025. [Online]. Available: https://eswd. eu/

[4] "National Centers for Environmental Information," *NCEI.* Accessed: Dec, 16, 2025. [Online]. Available: http://www.ncdc. noaa.gov/

[5] "Northern Tornadoes Project," *NTP.* Accessed: Dec, 16, 2025. [Online]. Available: https://uwo.ca/ntp/

[6] "Japan Meteorological Agency," *JMA.* Accessed: Dec, 16, 2025. [Online]. Available: https://www.data. jma.go.jp/stats/data/bosai/tornado/list. html

[7] B. Antonescu, D. M. Schultz, F. Lomas, and T. Kühne, "Tornadoes in Europe: Synthesis of the observational datasets," *Monthly Weather Review*, vol. 144, pp. 2445–2480, 2016. [Online]. Available: 10.1175/MWR-D-15-0298.1.

[8] A. Chernokulsky, M. Kurgansky, I. Mokhov, A. Shikhov, I. Azhigov, E. Selezneva, D. Zakharchenko, B. Antonescu, and T. Kühne, "Tornadoes in Northern Eurasia: From the Middle Age to the Information Era," *Monthly Weather Review*, vol. 148, pp. 3081–3110, 2020. [Online]. Available: 10.1175/MWR-D-19-0251.1.

[9] D. Veloso-Aguila, K. L. Rasmussen, and E. D. Maloney, "Tornadoes in Southeast South America: Mesoscale to Planetary-Scale Environments," *Monthly Weather Review*, vol. 152, pp. 295–318, 2024. [Online]. Available: 10.1175/MWR-D-22-0248.1.

[10] R. Kounkou, G. Mills, and B. Timbal, "A reanalysis climatology of cool-season tornado environments over southern Australia," *International Journal of Climatology*, vol. 29, pp. 2079–2090, 2009. [Online]. Available: 10.1002/joc.1856.

[11] H. Niino, T. Fujitani, and N. Watanabe, "A statistical study of tornadoes and waterspouts in Japan from 1961 to 1993," *Journal of Climate*, vol. 10, pp. 1730–1752, 1997. [Online]. Available: 10.1175/1520-0442(1997)010<1730:ASSOTA>2.0.CO;2.

[12] S. Kawazoe, M. Inatsu, M. Fujita, S. Sugimoto, Y. Okada, and S. Watanabe, "Evaluation of tornadic environments and their trends and projected changes in Japan," *npj Climate and Atmospheric Science*, vol. 6, 2023. [Online]. Available: 10.1038/s41612-023-00524-x.

[13] J. Chen, X. Cai, H. Wang, L. Kang, H. Zhang, Y. Song, H. Zhu, W. Zheng, and F. Li, "Tornado climatology of China," *International Journal of Climatology*, vol. 38, pp. 2478–2489, 2018. [Online]. Available: 10.1002/joc.5369.

[14] C. Zhang, M. Xue, K. Zhu, and X. Yu, "Climatology of significant tornadoes within China and comparison of tornado environments between the United States and China," *Monthly Weather Review*, vol. 151, pp. 465–484, 2023. [Online]. Available: 10.1175/MWR-D-22-0070.1.

[15] S. Bhan, S. Paul, K. Chakravarthy, R. Saxena, K. Ray, and N. Gopal, "Climatology of tornadoes over northwest India and Pakistan; and meteorological analysis of recent tornadoes over the region," *Journal of Indian Geophysical Union*, vol. 20, pp. 75–88, 2016.

[16] G. Capuli, "Project Severe Weather Archive of the Philippines (SWAP) Part 1: Establishing a baseline climatology for severe weather across the Philippine archipelago," *Annals of Geophysics*, vol. 67, Art. no. GC553, 2024. [Online]. Available: 10.4401/ag-9151.

[17] American Meteorological Society, "Tornado," 2024. [Online]. Available: https://glossary.ametsoc.org/wiki/Tornado.

[18] J. Rauhala, H. E. Brooks, and D. M. Schultz, "Tornado climatology of Finland," *Monthly Weather Review*, vol. 140, pp. 1446–1456, 2012. . [Online]. Available: 10.1175/MWR-D-11-00196.1.

[19] Soemantri, H. "Strategi Pengarasutamaan Pengurangan Resiko Bencana Di Sekolah". Jakarta, 2012.

[20] Supriyono, P. "Seri Pendidikan Pengurangan Risiko Bencana Angin Puting Beliung". Yogyakarta: Andi, 2015.

[21] International Energy Agency (IEA). "Technology Roadmap: Wind Energy -2013 edition".*IEA, 2013,* 2001.

[22] Burton, T., Nick J., David S., and Ervin B. "Wind Energy Handbook, 2nd Edition". *John Wiley & Sons, Ltd., West Sussex, U.K*, 2001.

[23] I. Ayundari & Sutikno, "Penentuan Zona Musim di Mojokerto Menurut Karakteristik Curah Hujan Dengan Metode Time Series Based Clustering," *INFERENSI*, vol. 2, no. 2, pp. 63–70, 2019.

[24] T. W. Liao, "Clustering of time series data—a survey," *Pattern Recognit*, vol. 38, no. 11, pp. 1857–1874, 2005.

[25] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, "Time Series Analysis : Forecasting and Control Fourth Edition". *New Jersey : John Wiley and Sons Inc.,* 2008. [Online]. Available: 10.1142/9789813148963_0009.
[26] A. Sardá-Espinosa, "Time-series clustering in R Using the dtwclust package," *R Journal*, vol. 11, no. 1, pp. 1–45, 2019, doi: 10.32614/rj-2019-023.
[27] H. Li, J. Liu, Z. Yang, R. W. Liu, K. Wu, and Y. Wan, "Adaptively constrained dynamic time warping for time series classification and clustering," *Inf Sci (N Y)*, vol. 534, pp. 97–116, 2020.
[28] M. A. A. Riyadi, D. S. Pratiwi, A. R. Irawan, and K. Fithriasari, "Clustering stationary and non-stationary time series based on autocorrelation distance of hierarchical and k-means algorithms," *International Journal of Advances in Intelligent Informatics*, vol. 3, no. 3, pp. 154–160, 2017.
[29] T. S. Madhulata, "An Overview of Clustering Methods," *IOSR Journal of Engineering*, vol. 2, no. 4, pp. 719–725, 2012, doi: 10.3233/ida-2007-11602.
[30] S. A. Nurfatimah, S. Hasna, dan D. Rostika, "Membangun Kualitas Pendidikan di Indonesia dalam Mewujudkan Program Sustainable Development Goals (SDGs)", *Jurnal Basicedu*, vol. 6, no. 4, pp. 6145–6154, 2022, [Online]. Available: 10.31004/basicedu.v6i4.3183
[31] Alwi, Wahidah, & Muh. Hasrul., "Analisis Klaster untuk Pengelompokan Kabupaten/Kota di Propinsi Sulawesi Selatan Berdasarkan Indikator Kesejahteraan Masyarakat". *Jurnal MSA*,Vol.6, No. 1, pp. 35-42, 2018.
[32] Ningsih, Silvia., Sri Wahyuningsih, & Yuki Novia Nasution., "Perbandingan Kinerja Metode Complete Linkage dan Average Linkage dalam Menentukan Analisis Cluster (Studi Kasus : Produksi Palawiya Provinsi Kalimantan Timur 2014/2015)". *Prosiding Seminar Sains dan Teknologi UNMUL*, Vol. 1 thn 2016, 46-50: Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Mulawarman, Samarinda, 2016.
[33] R. A. Johnson and D. W. Wichern, "Applied Multivariate Statistical Analysis". *Pearson Education, Inc.,* 2007.
[34] R. M. Shukla and S. Sengupta, "Scalable and robust outlier detector using hierarchical clustering and long short-term memory (lstm) neural network for the internet of things," *Internet of Things*, vol. 9, p. 100167, 2020.
[35] Y. Shi, "An Optimization Algorithm Based on Brainstorming Process," in *Emerging Research on Swarm Intelligence and Algorithm Optimization*, Y. Shi, Ed. Hershey, PA, *USA: IGI Global Scientific Publishing*, pp. 1–35, 2015. [Online]. Available: https://doi.org/10.4018/978-1-4666-6328-2.ch001
[36] Sanger, R. J., Fibriani, C., & Nataliani, Y., "Perancangan Aplikasi Sistem Informasi Pemantauan Kecepatan Angin Beserta Pengkategorian Jenis Angin dengan Hardware Inframerah Sebagai Media Kalibrasi", *Jurnal Teknologi Informasi – Aiti*, vol. 9, pp. 101–200, 2012. [Online]. Available: https://api.semanticscholar.org/CorpusID:116621043.
[37] Azhar, A., & Hashim, H., "A Review of Wind Clustering Methods Based on the Wind Speed and Trend in Malaysia," *Energies*, 16(8), 1–24, 2023. [Online]. Available: https://doi.org/10.3390/en16083388
[38] P. J. M. Ali & R. H. Faraj, "Data Normalization and Standardization: A Technical Report," *Machine Learning Technical Reports*, vol. 1, no. 1, pp. 1–6, 2014. [Online]. Available: doi:10.13140/RG.2.2.28948.04489
[39] C. S. Möller-Levet, F. Klawonn, K. H. Cho, and O. Wolkenhauer, "Fuzzy clustering of short time-series and unevenly distributed sampling points," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 2810, no. 0, pp. 330–340, 2003, doi: 10.1007/978-3-540-45231- 7_31.
[40] P. Galeano and D. Pella, "Multivariate Analysis in Vector Time Series," *Resenhas, the Journal of the Institute of Mathematics and Statistics of the University of Sao Paolo*, vol. 4, pp. 383–403, 2000.
[41] B. S. Everitt, S. Landau, M. Leese, & D. Stahl, "Cluster Analysis: Fifth Edition." *United Kingdom: John Wiley and Son, Ltd.*, 2011.
[42] Oti, Eric U., Olusola, Michael O, "Overview of Agglomerative Hierarchical Clustering Methods," *British Journal of Computer, Networking and Information Technology*, vol. 7, pp. 14-23, 2024. [Online]. Available: https://doi.org/10.52589/bjcnit-cv9poogw
[43] Chen, J., Wang, J., Cheng, S., Shi, Y. "Brain Storm Optimization with Agglomerative Hierarchical Clustering Analysis. In: Tan, Y., Shi, Y., Li, L. (eds) Advances in Swarm Intelligence." *ICSI 2016. Lecture Notes in Computer Science*(), vol 9713. Springer, Cham, 2016. [Online]. Available: https://doi.org/10.1007/978-3-319-41009-8_12
[44] Suwarti, Mulyono, & Prasetiyo, B., "Pembuatan Monitoring Kecepatan Angin dan Arah Angin Menggunakan Mikrokontroler Arduino", *Seminar Nasional Pendidikan, Sains dan Teknologi*, vol. 5, no. 1, pp. 56–64, 2017. [Online]. Available: https://jurnal.unimus.ac.id/index.php/psn12012010/article/viewFile/3152/3048

[45] F. Zhao, X. Hu, L. Wang, J. Zhao, J. Tang, and Jonrinaldi, "A reinforcement learning brain storm optimization algorithm with learning mechanism," *Knowledge-Based Systems*, vol. 235, 2021. [Online]. Available: https://doi.org/10.1016/j.knosys.2021.107645

[46] D. A. I. C. Dewi & D. A. K. Pramita, "Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali," *Matrix: Jurnal Manajemen Teknologi dan Informatika*, vol. 9, no. 3, pp. 102–109, 2019, [Online]. Available: 10.31940/matrix.v9i3.1662

[47] A. Mario, S. Herry, and H. Nasution, "Pemilihan Distance Measure Pada K-Means Clustering Untuk Pengelompokkan Member Di Alvaro Fitness," *Jurnal Sistem dan Teknologi Informasi*, vol. 1, no. 1, pp. 1–6, 2016.

[48] L. Kaufman dan P. J. Rousseeuw, "Finding Groups in Data: An Introduction to Cluster Analysis", 1991. [Online]. Available: 10.2307/2532178.

[49] "Average Wind Speed," *Badan Meteorologi, Klimatologi, dan Geofisika*. Accessed: Nov, 1, 2025. [Online]. Available: https://dataonline.bmkg.go.id/beranda.

[50] Riasa, Nyoman and Hartati, Rukmi and Manuaba, I B G and Santiari Dewa, "Pengaruh PLTB Sidrap Terhadap Sistem Kelistrikan Sulawesi Selatan," *Majalah Ilmiah Teknologi Elektro*, vol. 19, pp: 27, 2020. [Online]. Available: 10.24843/MITE.2020.v19i01.P04

[51] WAISNAWA, I Gd N Suta; RAJENDRA, I Made; SUDANA, I Made. " ANALISIS RISIKO INVESTASI PEMBANGKIT LISTRIK TENAGA BAYU DI NUSA PENIDA." *Matrix : Jurnal Manajemen Teknologi dan Informatika*, [S.l.], v. 5, n. 2, p. 1, feb. 2017. ISSN 2580-5630. [Online]. Available: https://ojs.pnb.ac.id/index.php/matrix/article/view/85

[52] A. Zakaria, I. Akbar, E. Muljadi, Suwarno and N. Hariyanto, "Grid Impact Study of Tanah Laut Wind Power Generation in Indonesia's Energy Transition," *2024 IEEE Industry Applications Society Annual Meeting (IAS)*, Phoenix, AZ, USA, 2024, pp. 1-6, [Online]. Available: 10.1109/IAS55788.2024.11023687

[53] A. Wulan andi, "TRANSISI ENERGI DI DAERAH TERPENCIL: STUDI KASUS IMPLEMENTASI SISTEM HYBRID PLTS-PLTD DI TEMAJOK, KALIMANTAN BARAT," *Technopex,* vol. 9, no. 1, pp. 270–276, 2025. [Online]. Available: https://sennas.iti.ac.id/tpx/index.php/c/article/view/103

[54] M. I. Amba and R. Dalimi, "Economic Analysis of Hybrid Power Plant (Solar-Diesel) on Kawaluso Island, North Sulawesi," *Jurnal EECCIS (Electrics, Electronics, Communications, Controls, Informatics, Systems),* vol. 17, no. 1, pp. 13–21, 2023. [Online]. Available: https://doi.org/10.21776/jeeccis.v17i1.1633